

## University of New Hampshire University of New Hampshire Scholars' Repository

---

Center for Coastal and Ocean Mapping

Center for Coastal and Ocean Mapping

---

8-2002

# Underwater Video Survey: Planning and Data Processing

Yuri Rzhanov

*University of New Hampshire, Durham, [Yuri.Rzhanov@unh.edu](mailto:Yuri.Rzhanov@unh.edu)*

Lloyd C. Huff

*University of New Hampshire, Durham*

Randy G. Cutter Jr.

*University of New Hampshire, Durham*

Follow this and additional works at: <https://scholars.unh.edu/ccom>



Part of the [Computer Sciences Commons](#), and the [Oceanography and Atmospheric Sciences and Meteorology Commons](#)

---

### Recommended Citation

Rzhanov, Yuri; Huff, Lloyd C.; and Cutter, Randy G. Jr., "Underwater Video Survey: Planning and Data Processing" (2002). *International Conference on Signal and Image Processing (SIP)*. 254.  
<https://scholars.unh.edu/ccom/254>

This Conference Proceeding is brought to you for free and open access by the Center for Coastal and Ocean Mapping at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Center for Coastal and Ocean Mapping by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact [nicole.hentz@unh.edu](mailto:nicole.hentz@unh.edu).

# UNDERWATER VIDEO SURVEY: PLANNING AND DATA PROCESSING

*Y. Rzhanov*

Center for Coastal and Ocean Mapping (C-COM)  
University of New Hampshire  
Durham 03824, USA  
email: yuri.rzhanov@unh.edu

## ABSTRACT

The importance of underwater video surveys as an exploration tool has been steadily increasing over recent years [1]. Better photographic equipment, more effective sources of illumination, and improved processing techniques - all make video surveying a reliable tool for seafloor habitat mapping, sediment boundary delineation and groundtruthing, mapping and documentation of forensic and archaeological sites. There is a change in attitude towards video surveying that affects the way the data is collected, and hence its quality. Earlier video data processing algorithms had to cope with whatever was recorded (often simultaneously with acquisition of other data, considered to be more important). Now we have a chance to plan ahead and organize a survey in a way most suitable for the processing.

The goal of this paper is to review available processing techniques and to discuss preferable survey patterns, associated errors and processing stability.

## 1. INTRODUCTION

For groundtruthing, video surveys are usually conducted in straight lines. The camera is attached to a towbody or other underwater vehicle [marks,singh], or is held by a diver [icip00]. Geo-referencing of the imagery is relatively poor, accuracy is typically lower than 3-5 meters (poor in comparison with the sub-centimeter resolution of the images), but is sufficient for the purpose - determination of boundaries between sediments/habitats. The width of a mosaic created from these images depends on required resolution and the visibility range in the water (given the source of illumination). Mosaics guarantee accurate recognition of benthic species (or geological formations) and allow estimation of their relative densities [5].

In such a "linear survey", taken from a relatively stable platform (special thanks to Bobby Forbes who proved that a diver can be a stable platform!), perspective distortions are minimal and we found that the most successful registration of frames can be done with a featureless frequency domain-

based technique (FFD) discussed in [6, 4]. Although it allows recovery of only 4 parameters of the affine transformation relating two consecutive images (rotation and three translations), the advantages in comparison with other techniques are apparent. Optimization-based methods rely on the brightness constancy constraint and become unreliable when artificial illumination is necessary. Feature-matching performs poorly in the absence of well-defined features and is error-prone when something distinct is moving in a field of view (both situations are quite common in underwater imagery).

The FFD method is tolerant to illumination inhomogeneities and shadows (in fact, gradual darkening of the image towards the edges improves reliability) and ignores small objects moving in a frame (in processing they manifest as a non-coherent noise or as false peaks that are easy to discriminate). This inherent robustness makes the FFD method exceptionally useful, even if perspective distortions are significant and co-registration transformation cannot be limited to affine. Another type of survey is required when the area of interest is significantly larger than the camera field of view. The final product of such a survey is a map of the area - projection of the features onto some chosen planar surface. If the terrain is slanted, the map could be created in two different ways: a) the camera stays at a constant height and looks vertically down, or b) the camera is at a constant distance from the terrain and looks normal to the surface. The better way is a matter of preference.

In the scope of this paper we consider a somewhat simplified case of a single-swath survey, with the platform moving in a straight line, but with the increased swath width by modulation of the camera tilt [5]. Pitch and roll are changing periodically; coverage pattern depends on the amplitudes and phases of these two motions. (Assume that imagery taken at oblique angles provides sufficient resolution and that the ambient light or artificial illumination is also sufficient.)

Overlap between non-consecutive frames in an acquired sequence allows for co-registration of these frames and this additional information makes a final mosaic less affected by

the accumulation of errors.

In the previous paper[7] the apparatus used for shallow water video surveying was described in detail. Imagery is acquired by a consumer-grade Sony digital video camera, connected to a single-board computer (Jackrabbit, model BL1800) through a Control-L device. This allows time-code recording by the computer synchronously with the input from a tilt/compass sensor (Precision Navigation, model TCM2-50) attached to the camera and a GPS receiver.

## 2. VIDEO DATA PROCESSING

The processing stages of the collected imagery are as follows (a pre-processing stage includes correction of lens distortion and possible application of a mask removing influence of permanent occlusions).

1. Consecutive frames corresponding to the records from a log file are co-registered. Note that due to significant (and essential for our purpose) tilt of the camera, the transformation model cannot be restricted to a 4-parameter affine. The 8-parameter perspective model was employed, and finding the transformation for a pair of frames involves application of an optimization algorithm. As mentioned above, it is based on the brightness constancy constraint, and its success depends on image quality, scene illumination, and an initial guess supplied to the algorithm. Deep-water surveys need artificial illumination which cause brightness pattern moving with the platform (and changing as the height over the seabed is changing), and hence requires removal of this pattern (by de-trending [4], for example) prior to optimization.

In difficult cases with highly nonuniform lighting the pre-processing may require adaptive histogram equalization and/or edge detection, to keep only those features that do not depend on illumination [2].

Finding a suitable initial guess for optimization sometimes poses a non-trivial problem. When obvious candidates for the initial guess (unit transform, successful result of co-registration for previous or next pair of frames, etc.) do not lead to the correct registration, recorded values of the camera tilt can be used to re-project frames onto a common surface. Next they are robustly co-registered using the FFD method, and parameters of the found affine transformation are used in conjunction with the camera tilt to calculate a good initial guess for the 8-parameter optimization process.

Pairwise co-registration produces a number of important estimates, that may be used during further processing: frames' overlap, average per-pixel error, local areas of high error, etc. Second derivatives of the

penalty function with respect to model parameters indicate relative importance of the components of the found solution: a large 2-nd derivative signifies that variation of this variable should be heavily penalized; a small 2-nd derivative suggests that the value may be easily altered if other constraints are considered.

2. Choosing a global re-projection surface (e.g., flat horizontal) and world (global) transformation for the initial frame in a sequence, world transformations for all following frames are calculated chain-like. Errors grow monotonously - due to various reasons among which the most important is the limited accuracy of co-registration transformations found in the previous stage.
3. World transformation for the frames and information from other sources are used in a global alignment procedure - to minimize accumulated errors and create a seamless mosaic with maximum possible accuracy. Rather than formulating a variant of "bundle adjustment" algorithm (see, for example, [8]), a simplified approach for global alignment of a sequence of images related by affine transformations, suggested in [6], is followed. Co-registration transformations between all overlapping non-sequential frames are calculated (this procedure is reliable and fast as a reasonable initial guess for the optimization algorithm is known). Then a sparse linear system of equations is built where world transformations for images are treated as unknowns. With non-sequential elements added, the system is over-constrained and can be solved only in a least squares sense. Note that frames as such do not participate in the calculations; however important information about them is used in the least squares problem as weighting coefficients: overlap between two frames reflects reliability of the transformation estimate, and 2-nd derivatives array mentioned above - relative importance of the particular component. Non-sequential overlap is conveniently expressed in terms of a "support matrix". Its structure reflects the survey pattern: diagonal elements are always 1's (each frame fully supports itself); width of the diagonal "belt" reflects speed of the camera motion; off-diagonal clusters correspond to the camera looping back to the areas already covered.

Note that information from the sensors recorded synchronously with the timecode can be incorporated in the least squares problem - with the weighting coefficients reflecting accuracy of the corresponding measurement.

Case	$A_P$	$A_R$	$P_P$	$P_R$	$\phi_P$	$\phi_R$
A	20	0	1.5	1.0	0	0
B	20	22.5	1.0	2.0	0	0
C	20	20	1.0	1.0	0	0
D	20	20	1.0	0.75	0	0.2
E	20	15	2.0	1.0	0	-0.25
F	20	22.5	1.0	1.5	0	0
G	20	15	1.5	0.75	-0.5	0
H	20	17.5	1.0	1.0	0	-0.25

Table 1: Coefficients describing camera pitch and roll.

### 3. SURVEY PLANNING AND SIMULATION RESULTS

Relative accuracy and stability of video survey patters were investigated in a numerical simulation of video acquisition process. Processing algorithm outlined above was then applied to acquired imagery. Camera motion was estimated from the calculated world transformations and the results were compared with known true values.

8 different patterns were chosen. Camera pitch and roll are described by the following formulas:

$$\begin{aligned} \text{Pitch} &= A_P \sin(2\pi(\frac{t}{P_P} + \phi_P)), \\ \text{Roll} &= A_R \sin(2\pi(\frac{t}{P_R} + \phi_R)), \end{aligned}$$

where  $A_P$  and  $A_R$  are the amplitudes,  $P_P$  and  $P_R$  - periods, and  $\phi_P$  and  $\phi_R$  - phases of the motions, respectively. Numerical values for the 8 cases are presented in Table 1. In all cases the camera is moving with a constant speed along the X axis (see Figure 1), and all sequences contain 50 frames. Coefficients were chosen such that each survey covers approximately the same area of the imaged surface.

Survey patterns are easy to visualize by plotting centers of the re-projected frames (Figure 1).

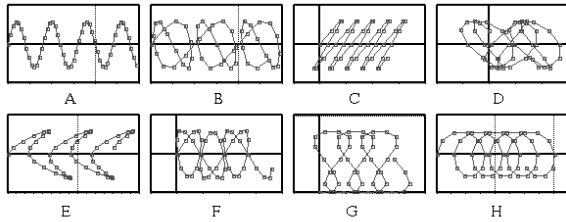


Figure 1: Centers of re-projected frames for 8 survey patterns.

Co-registration of consecutive frames in the sequences allows generation of the corresponding support matrices, shown in Figure 2 as brightness images: white means full support, black - no overlap.

Once world transformations for all frames were found, the model described in [9] was used to estimate correspond-

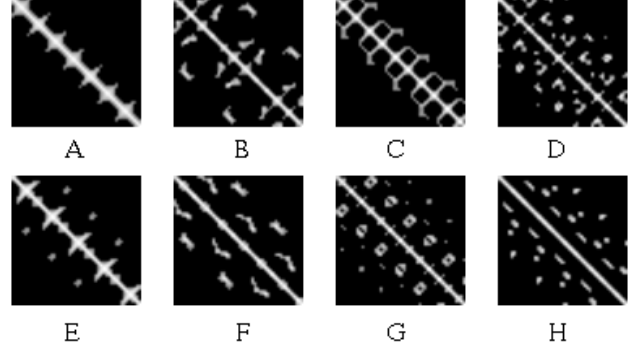


Figure 2: Support matrices for the above survey patterns.

Case	Before GA		After GA		Overlap	
	Distance	Angle	Distance	Angle	Avg	Min
A	1.03	0.42	0.61	0.24	80	71
B	0.71	0.29	0.64	0.20	65	44
C	1.54	0.77	0.85	0.26	61	38
D	2.71	1.03	0.91	0.33	53	30
E	1.02	0.55	1.00	0.64	74	58
F	1.31	0.58	1.12	0.54	66	46
G	2.04	1.39	1.28	0.52	59	46
H	1.82	0.72	1.35	0.44	58	49

Table 2: Average difference between computed and groundtruth values for camera position (normalized) and orientation (degrees), before and after the global alignment (GA) procedure; average ("Avg") and minimum ("Min") overlap between consecutive frames (percent).

ing camera position and orientation. These values (before and after global alignment) are compared with the known "true" values. The comparison results are summarized in Table 3.

The analysis shows that for some patterns (specifically, E and F) the global alignment step does not improve the overall picture. In the other cases the accumulated error is not significant even prior to the global optimization, but the optimization can improve the alignment even further. Clearly, the above choice of patterns is not exhaustive and the error analysis is phenomenological. However some observations can be made. Low errors in case A (simple "shevron"-type pattern) can be attributed to a high level of overlap between consecutive frames and hence accurate determination of relative transformations between them. Similarly, low overlaps in cases F, G and H may be responsible for high error. Cases B and C, on the contrary, combine relatively low overlap with accurate reconstruction of the camera path and resulting mosaic. These survey patterns are likely to provide better results in video surveying.

#### 4. CONCLUSIONS

Possible improvement of video mosaic creation by means of survey planning was discussed. Suggestions were verified by numerical experiments. The apparatus used for video data acquisition and processing stages are reviewed and discussed.

#### 5. REFERENCES

- [1] K. Lebart, Y. Petillot, C. Smith, D. Lane, E. Trucco, "Video sensors play major role in subsea scientific missions", *Sea Technology*, February 2002, pp. 10-14.
- [2] R. L. Marks, S. M. Rock, M. J. Lee, "Real-time video mosaicking of the ocean floor." *IEEE Journal of Oceanic Engineering*, vol. 20, pp. 229-241, 1995.
- [3] H. Singh, L. Whitcomb, D. Yoerger, O. Pizarro, "Microbathymetric mapping from underwater vehicles in the deep ocean." *Computer Vision and Image Understanding*, vol. 79, pp. 143-161, 2000.
- [4] Y. Rzhanov, L. M. Linnett, R. Forbes, "Underwater mosaicing for seabed mapping." 2000 International conference on Image Processing, ICIP2000, Vancouver, Canada, vol. 1, pp. 224-227, 2000.
- [5] Y. Rzhanov, G. R. Cutter, L. Huff, "Sensor-assisted video mosaicing for seafloor mapping." 2001 International conference on Image Processing, ICIP2001, Thessaloniki, Greece, vol. 2, pp. 411-414, 2001.
- [6] J. Davis, 1998, *Mosaics of scenes with moving objects*. IEEE Comput. Soc. Proceedings, vol. 1, pp. 354-360.
- [7] Y. Rzhanov, L. Huff, G. R. Cutter, "Improvement of image alignment using camera attitude information", *Sixth International Symposium On Signal Processing and Its Applications*, Kuala Lumpur, Malaysia, 2001
- [8] B. Triggs, P. McLauchlan, R. Hartley, A. Fitzgibbon, "Bundle adjustment - a modern synthesis", *Vision Algorithms: Theory and Practice*, B. Triggs, A. Zisserman and R. Szeliski (Eds.) Springer-Verlag LNCS, 2000.
- [9] Y. Rzhanov, L. Huff, G. R. Cutter, "Seafloor video mapping: Modeling, algorithms, apparatus", Submitted to 2002 International conference on Image Processing, ICIP2002, New-York, USA.